



DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Optimum-Weighted RLS Channel Estimation for Rapid Fading MIMO Channels

The Harvard community has made this article openly available.

[Please share](#) how this access benefits you. Your story matters.

Citation	Toshiaki, Koike-Akino. 2008. Optimum-weighted RLS channel estimation for rapid fading MIMO channels. IEEE Transactions on Wireless Communications 7(11): 4248-4260.
Published Version	http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=7693
Accessed	February 17, 2015 4:24:50 PM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:2748553
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

(Article begins on next page)

Optimum-Weighted RLS Channel Estimation for Rapid Fading MIMO Channels

Toshiaki Koike Akino, *Member, IEEE*

Abstract—This paper investigates on an accurate channel estimation scheme for fast fading channels in multiple-input multiple-output (MIMO) mobile communications. A high-order exponential-weighted recursive least-squares (EW-RLS) method has been known as a good channel estimation scheme in rapid fading. However, there exists a drawback that we need to properly adjust the estimation parameters of a forgetting factor and an estimation order according to the channel environment. In this paper, we theoretically derive an optimum-weighted LS (OW-LS) channel estimation based on the statistical knowledge of the spatio-temporal channel correlation. Through the analysis, we reveal that the zero-th order polynomial becomes optimal when the optimum-weighting is employed. Furthermore, we propose an efficient recursive algorithm for channel tracking in order to reduce the computational complexity. Since the proposed scheme automatically adapts the weighting coefficients to the channel condition, it has a significant advantage in mean-square error (MSE) performance compared to the EW-RLS scheme.

Index Terms—Weighted least-squares (LS) channel estimation, multiple-input multiple-output (MIMO), optimum-weighting, recursive algorithm, fast fading channel.

I. INTRODUCTION

IN the past several years, radio communications systems exploiting multiple antennas at both transmitter and receiver, known as multiple-input multiple-output (MIMO) systems, have received a significant amount of attention. The MIMO systems have a potential to achieve a substantial increase of the spectral efficiency under rich-scattering wireless channels [2, 3]. Therefore, it has been a promising technique to realize Gbps-class high-speed transmissions for near-future mobile communications under slow fading channels [4, 5].

An accurate channel estimation has been one of the most important issues for reliable wireless communications. In mobile communications, we should generally deal with a time-varying fading channel resulting from the receiver's traveling and the surrounding scatters' movements. To combat the channel dynamics, an exponential-weighted least-squares (EW-LS) channel estimation [6, 7] is commonly used since it offers a good estimation. It is based on the Kalman filtering, and implemented as the recursive LS (RLS) algorithm to avoid computing an inversion matrix directly. In the EW-RLS estimation scheme, the so-called forgetting factor is used to overcome the channel variation. The channel estimation error

is severely dependent on the forgetting factor; specifically, the larger value decreases the error caused by the additive noise, whereas the smaller value decreases the error caused by the channel variation. However, the optimum forgetting factor exactly minimizing the total channel estimation error has not been derived in a theoretical manner so far (although we can find an effort optimizing a forgetting factor by using a frequency domain approach in [8], the derived solution is not optimal). Correspondingly, we should properly adjust it according to the channel environments in a numerical, experimental or manual manner. After the adjustment, a fixed value of the forgetting factor is typically used in a general application. Although several modified schemes have been studied on employing variable forgetting factor [9–14] and parallel forgetting factors [15], the channel estimation accuracy is severely degraded in an extremely fast fading channel even when the forgetting factor is well optimized.

In such a fast fading channel, an improved channel modeling is effective for channel estimation: e.g., an autoregressive (AR) process model [16–18] and a high-order polynomial channel model [19, 20]. We refer to the latter technique applying the polynomial modeling to the LS estimation as the high-order LS estimation. Since the high-order LS estimation offers a good performance in channel estimation under rapid fading environments, we investigate on it in this paper. The high-order EW-RLS scheme is one of regression methods based on a polynomial fitting with exponential weights. The higher-order polynomial can improve the robustness against the channel dynamics, whereas it can increase the channel estimation error caused by the additive noise because the number of parameters to be estimated is increased [20]. Hence, there exists a drawback that we need to adjust the polynomial order as well as the forgetting factor. To overcome the degradation due to their mis-adjustments, we have proposed an algorithm diversity scheme in [15], in which multiple EW-RLS channel estimations with different order and different forgetting factor are employed in parallel for a practical use. As an alternative solution, in this paper, we theoretically optimize the weighting coefficients (not the forgetting factor in exponential weighting) and the polynomial order in the high-order weighted LS channel estimation scheme for time-varying MIMO fading channels. Here, we focus on deriving an exact optimum solution for arbitrary-weighting coefficients. To the best of the author's knowledge, there is no investigation on such an optimum-weighted LS (OW-LS) channel estimation for MIMO systems, although some related literatures for single-antenna systems can be found in [8, 21]. In addition to the extension to the multiple-antenna case from the single-antenna

Manuscript received May 14, 2007; revised August 21, 2007 and October 23, 2007; accepted January 27, 2008. The associate editor coordinating the review of this paper and approving it for publication was M. Saquib. This paper has been partly reported in *Proc. IEEE ICC 2007* [1].

The author is with the School of Engineering and Applied Sciences, Harvard University, 33 Oxford Street, Cambridge, MA 02138, USA (e-mail: koike@seas.harvard.edu).

Digital Object Identifier 10.1109/T-WC.2008.070497

case studied in [8, 21], our contributions include a theoretical optimization of the polynomial order for the high-order LS, a consideration of more general channel dynamics, and a derivation of the recursive tracking algorithm.

At first, we theoretically derive an optimum weighting, that exactly minimizes a mean-square error (MSE) of the estimated channel matrix. For the derivation, a spatio-temporal correlation of the MIMO channel matrix is dealt with for a realistic application. Next, we show a theoretical verification of the interesting fact that the higher-order polynomial channel model is not advantageous when we can optimize the weighting coefficients. Then, we develop a computationally efficient recursion technique called as an optimum-weighted RLS (OW-RLS) algorithm, for decision-directed channel tracking. At last, we demonstrate the advantage of the proposed scheme through performance evaluations with a comparison to the other schemes. The main contributions of this paper are threefold as summarized below:

- we derive theoretically optimal weighting for high-order weighted LS in MIMO fading channels,
- we prove that the 0-th order is the best polynomial order for optimum weighting,
- we propose a recursive algorithm for decision-directed channel tracking to reduce the computational complexity.

II. WEIGHTED LEAST-SQUARES

Notations: Throughout this paper, matrices and vectors are described by capital and lower case letters in boldface. Let \mathbf{X} and x be a complex matrix and a complex number, respectively. Notations $\|\mathbf{X}\|$, \mathbf{X}^T , \mathbf{X}^\dagger , \mathbf{X}^{-1} and $\text{tr}[\mathbf{X}]$ represent the Frobenius norm, the transpose, the Hermitian transpose, the inverse and the trace of a matrix \mathbf{X} , respectively. The magnitude and the complex conjugate of x are denoted as $|x|$ and x^* . The sets of complex numbers and real numbers are designated by \mathbb{C} and \mathbb{R} , respectively. The matrix \mathbf{I}_p denotes the p -dimensional identity matrix. We use $\mathbb{E}[\cdot]$ for the expectation function.

A. Signal Description

We consider a $M \times N$ MIMO multiplexing system, where M transmit antennas and N receive antennas are used. A transmission frame consists of a K_i -symbol information sequence preceded by a K_p -symbol pilot sequence for channel acquisition. We use a discrete-time and low-pass equivalent signal representation for a symbol rate of $1/T_s$. We suppose that the received signal at the k -th symbol is modeled as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k, \quad (1)$$

where $\mathbf{y}_k \in \mathbb{C}^{N \times 1}$, $\mathbf{H}_k \in \mathbb{C}^{N \times M}$, $\mathbf{x}_k \in \mathbb{C}^{M \times 1}$ and $\mathbf{w}_k \in \mathbb{C}^{N \times 1}$ represent the received signals vector, the MIMO channel matrix, the transmitted signals vector, and the additive Gaussian noise vector, respectively. We assume that the additive noises are independently distributed in both time and spatial dimensions as follows: $\mathbb{E}[\mathbf{w}_k \mathbf{w}_j^\dagger] = \sigma^2 \delta_{k-j} \mathbf{I}_N$, where σ^2 is a noise variance, and δ_i denotes the Kronecker's delta operator; i.e. $\delta_i = 1$ when $i = 0$, otherwise $\delta_i = 0$.

When the channel variation is sufficiently slow such that all channel matrices \mathbf{H}_k are nearly identical during the interested

frame, we should only estimate one common channel matrix $\mathbf{H}^{(0)} \simeq \mathbf{H}_k$ for all symbol index k . In fast fading, the channel matrix variation forms some sort of random curves as a function of the symbol index k . In such a time-varying channel with additive noises, we may use a high-order polynomial channel model [19, 20], assuming that the unknown curve of the channel variation is approximated to the D -th order Taylor series expansion according to the symbol instance k :

$$\mathbf{H}_k \simeq \sum_{d=0}^D k^d \mathbf{H}^{(d)} = \mathcal{H} \mathbf{Q}_k, \quad (2)$$

where $\mathbf{H}^{(d)} \in \mathbb{C}^{N \times M}$ denotes the d -th Taylor coefficient matrix, and expansion matrices \mathcal{H} and \mathbf{Q}_k are defined as

$$\mathcal{H} \triangleq [\mathbf{H}^{(0)}, \mathbf{H}^{(1)}, \dots, \mathbf{H}^{(D)}] \in \mathbb{C}^{N \times M(D+1)}, \quad (3)$$

$$\mathbf{Q}_k \triangleq [k^0 \mathbf{I}_M, k^1 \mathbf{I}_M, \dots, k^D \mathbf{I}_M]^T \in \mathbb{R}^{M(D+1) \times M}. \quad (4)$$

Supposed that we can use previously transmitted signals as a K -symbol training sequence at the K -th symbol instance, and we neglect any decision error in this sequence for simplicity (it will be relaxed in performance evaluation), we estimate an expanded channel matrix $\hat{\mathcal{H}}_{K+1}$ for detecting the $(K+1)$ -th symbol \mathbf{x}_{K+1} using the weighted LS method as follows:

$$\hat{\mathcal{H}}_{K+1} = \arg \min_{\hat{\mathcal{H}} \in \mathbb{C}^{N \times M(D+1)}} \sum_{k=1}^K \beta_k \left\| \mathbf{y}_k - \hat{\mathcal{H}} \mathbf{Q}_k \mathbf{x}_k \right\|^2, \quad (5)$$

where $\beta_k \in \mathbb{R}$ stands for a weighting coefficient. The LS solution is given by

$$\hat{\mathcal{H}}_{K+1} = \left(\sum_{k=1}^K \beta_k \mathbf{y}_k \mathbf{x}_k^\dagger \mathbf{Q}_k^T \right) \left(\sum_{k=1}^K \beta_k \mathbf{Q}_k \mathbf{x}_k \mathbf{x}_k^\dagger \mathbf{Q}_k^T \right)^{-1}. \quad (6)$$

Thus, we can obtain the estimated channel matrix:

$$\begin{aligned} \hat{\mathbf{H}}_{K+1} &= \hat{\mathcal{H}}_{K+1} \mathbf{Q}_{K+1} \\ &= \sum_{k=1}^K \mathbf{y}_k \mathbf{g}_k^\dagger = \mathbf{Y}_K \mathbf{G}_K^\dagger, \end{aligned} \quad (7)$$

where we defined

$$\mathbf{g}_k^\dagger \triangleq \beta_k \mathbf{x}_k^\dagger \mathbf{Q}_k^T \Phi_K^{-1} \mathbf{Q}_{K+1} \in \mathbb{C}^{1 \times M}, \quad (8)$$

$$\Phi_K \triangleq \sum_{k=1}^K \beta_k \mathbf{Q}_k \mathbf{x}_k \mathbf{x}_k^\dagger \mathbf{Q}_k^T \in \mathbb{C}^{M(D+1) \times M(D+1)}, \quad (9)$$

$$\mathbf{G}_K \triangleq [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_K] \in \mathbb{C}^{M \times K}, \quad (10)$$

$$\mathbf{Y}_K \triangleq [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K] \in \mathbb{C}^{N \times K}. \quad (11)$$

Under the definitions above, one can see that the matrix \mathbf{G}_K must have the following property:

$$\sum_{k=1}^K \mathbf{Q}_k \mathbf{x}_k \mathbf{g}_k^\dagger = \mathbf{X}_K \mathbf{G}_K^\dagger = \mathbf{Q}_{K+1}, \quad (12)$$

where \mathbf{X}_K is an expanded transmission signals matrix for the D -th order estimation, defined as

$$\mathbf{X}_K \triangleq [\mathbf{Q}_1 \mathbf{x}_1, \mathbf{Q}_2 \mathbf{x}_2, \dots, \mathbf{Q}_K \mathbf{x}_K] \in \mathbb{C}^{M(D+1) \times K}. \quad (13)$$

The higher-order LS can offer an accurate channel estimation in fast fading channels, compared to the 0-th order

one [19,20]. However, it may increase the computational complexity, and moreover, it has a negative effect in channel estimation accuracy at low signal-to-noise ratio (SNR) and in slow fading channels. It is because the number of channel parameters to be estimated is increased by DNM , i.e. $\mathbf{H}^{(d)}$ for $d = 1, 2, \dots, D$. As a result, the higher-order estimation has poorer convergence for initial channel acquisition, leading to a need of a longer pilot sequence for convergence. Nevertheless, since the high-order LS gives consideration to the high-order gradients of the channel variation, it is much advantageous for rapid fading channels when the polynomial order is properly chosen [20].

B. Exponential-Weighted RLS Channel Estimation

As shown in (6), the LS solution requires a matrix inversion in $\Phi_K^{-1} = (\sum_{k=1}^K \beta_k \mathbf{Q}_k \mathbf{x}_k \mathbf{x}_k^\dagger \mathbf{Q}_k^\dagger)^{-1}$, which generally needs a cubic complexity order according to the matrix size [22]. However, the computational complexity can be reduced to a square complexity order when we employ an exponential weighting, where the weighting coefficient is given by $\beta_k = \lambda^{K-k}$. The parameter $\lambda \in \mathbb{R}$ is referred to as a forgetting factor [6]. In the exponential-weighted LS, the covariance matrix Φ_K can be recursively written as $\Phi_K = \lambda \Phi_{K-1} + \mathbf{Q}_K \mathbf{x}_K \mathbf{x}_K^\dagger \mathbf{Q}_K^\dagger$. Using the Sherman-Morrison formula [22], its inversion is given by $\Phi_K^{-1} = \lambda^{-1} \Phi_{K-1}^{-1} - \lambda^{-1} \Phi_{K-1}^{-1} \mathbf{Q}_K \mathbf{x}_K \mathbf{x}_K^\dagger \mathbf{Q}_K^\dagger \Phi_{K-1}^{-1} / (\lambda + \mathbf{x}_K^\dagger \mathbf{Q}_K^\dagger \Phi_{K-1}^{-1} \mathbf{Q}_K \mathbf{x}_K)$ in a recursive way. This formula is used for implementing the following recursive algorithm [6], called as the D -th order EW-RLS algorithm:

$$\begin{aligned} \zeta_K &= \Phi_{K-1}^{-1} \mathbf{Q}_K \mathbf{x}_K \in \mathbb{C}^{M(D+1) \times 1}, \\ \alpha_K &= \frac{1}{\lambda + \zeta_K^\dagger \mathbf{Q}_K \mathbf{x}_K} \in \mathbb{R}, \\ \Phi_K^{-1} &= \lambda^{-1} \Phi_{K-1}^{-1} - \lambda^{-1} \alpha_K \zeta_K \zeta_K^\dagger \in \mathbb{C}^{M(D+1) \times M(D+1)}, \\ \mathbf{e}_K &= \mathbf{y}_K - \hat{\mathbf{H}}_K \mathbf{x}_K \in \mathbb{C}^{N \times 1}, \\ \hat{\mathbf{H}}_{K+1} &= \hat{\mathbf{H}}_K + \alpha_K \mathbf{e}_K \zeta_K^\dagger \in \mathbb{C}^{N \times M(D+1)}, \\ \hat{\mathbf{H}}_{K+1} &= \hat{\mathbf{H}}_{K+1} \mathbf{Q}_{K+1} \in \mathbb{C}^{N \times M}. \end{aligned}$$

For initialization, $\hat{\mathbf{H}}_1$ is often zeroed, and Φ_0^{-1} is set to $\rho \mathbf{I}_{M(D+1)}$, where $\rho \gg 1$ is a constant with a sufficiently large value. However, since the value of ρ can greatly influence a convergence behavior along with an arithmetic stability, as reported in [9], it is better to use a direct matrix inversion $\Phi_{K_p}^{-1}$ at the end of the pilot sequence. Subsequently, the RLS algorithm should be used for channel tracking in the information period. In particular, the higher-order EW-RLS algorithm becomes instable when we use a fixed initial value of ρ . The receiver can compute the inversion matrix $\Phi_{K_p}^{-1}$ in advance, because it depends only on the forgetting factor and the pre-determined pilot sequence.

Introducing the forgetting factor contributes to not only reducing the complexity but also combating a channel dynamics, because it gives more importance to the recently received signal information with an exponential weight. However, the channel estimation accuracy is severely dependent on the forgetting factor [6]. It is well-known that a large value of the forgetting factor can decrease the error caused by the

additive noise (often called as the noise error), whereas a small value is required to decrease the error due to the channel variation (often called as the lag error). Therefore, we should properly set the forgetting factor depending on the channel environment in order to minimize the total channel estimation error. Since the optimum value of the forgetting factor has not been derived in a theoretical manner, a fixed value is typically used after evaluating the estimation accuracy in a numerical or experimental sense. In [8], a sub-optimal forgetting factor for the 0-th order EW-RLS has been reported for Rayleigh fading channels in single antenna systems; more specifically, $\lambda_{\text{opt}} \simeq 1 - (2(2\pi f_D T_s)^2 / \sigma^2)^{1/3}$ where f_D denotes the maximum Doppler frequency. However, it cannot be used in direct for the high-order EW-RLS channel estimation in MIMO fading channels. In this paper, we optimize the weighting coefficients $\{\beta_k\}$ instead of optimizing the forgetting factor λ , and theoretically derive an optimum-weighted (OW) LS channel estimation scheme which adaptively adjusts the weighting coefficients according to the channel environment. As is obvious, the OW-LS always outperforms the EW-RLS even when the forgetting factor is well optimized or adapted by [9–14].

C. Mean-Square Error of Weighted LS Channel Estimation

Before deriving the OW-LS channel estimation scheme, we show MSE of estimated channel matrix, dealing with a spatio-temporal correlation. The MSE of the estimated channel matrix at the target symbol ($K+1$) is defined as:

$$\begin{aligned} \varepsilon_{K+1} &\triangleq \mathbb{E} \left[\left\| \mathbf{H}_{K+1} - \hat{\mathbf{H}}_{K+1} \right\|^2 \right] \\ &= \text{tr} \left[\mathbb{E} \left[\mathbf{H}_{K+1}^\dagger \mathbf{H}_{K+1} + \mathbf{G}_K \mathbf{Y}_K^\dagger \mathbf{Y}_K \mathbf{G}_K^\dagger \right. \right. \\ &\quad \left. \left. - \mathbf{G}_K \mathbf{Y}_K^\dagger \mathbf{H}_{K+1} - \mathbf{H}_{K+1}^\dagger \mathbf{Y}_K \mathbf{G}_K^\dagger \right] \right]. \end{aligned} \quad (14)$$

Here, we assume that the MIMO channel has a spatio-temporal correlation written as

$$\mathbb{E} \left[\mathbf{H}_k^\dagger \mathbf{H}_j \right] = \boldsymbol{\Theta}_{k-j} \in \mathbb{C}^{M \times M}. \quad (15)$$

Notice that when there is no spatial correlation, the correlation matrix becomes $\boldsymbol{\Theta}_{k-j} = N \rho_{k-j} \mathbf{I}_M$, where ρ_{k-j} is a temporal correlation coefficient. When the channel is richly scattered following the Jakes model [23], we get $\rho_{k-j} = J_0(2\pi f_D T_s(k-j))$, where $J_0(\cdot)$ denotes the zero-th order of the first kind Bessel function.

Making use of the correlation matrices $\{\boldsymbol{\Theta}_i\}$, we have

$$\begin{aligned} \mathbb{E} \left[[\mathbf{Y}_K^\dagger \mathbf{Y}_K]_{i,j} \right] &= E \left[\mathbf{y}_i^\dagger \mathbf{y}_j \right] \\ &= E \left[(\mathbf{H}_i \mathbf{x}_i + \mathbf{w}_i)^\dagger (\mathbf{H}_j \mathbf{x}_j + \mathbf{w}_j) \right] \\ &= \mathbf{x}_i^\dagger \boldsymbol{\Theta}_{i-j} \mathbf{x}_j + N \sigma^2 \delta_{i-j}, \end{aligned} \quad (16)$$

$$\begin{aligned} \mathbb{E} \left[[\mathbf{Y}_K^\dagger \mathbf{H}_{K+1}]_{i,\cdot} \right] &= E \left[\mathbf{y}_i^\dagger \mathbf{H}_{K+1} \right] \\ &= E \left[(\mathbf{H}_i \mathbf{x}_i + \mathbf{w}_i)^\dagger \mathbf{H}_{K+1} \right] \\ &= \mathbf{x}_i^\dagger \boldsymbol{\Theta}_{i-(K+1)}, \end{aligned} \quad (17)$$

where $[\mathbf{A}]_{i,j}$ and $[\mathbf{A}]_{k,\cdot}$ denote, respectively, the i -th row and j -th column entry, and the k -th row vector of a matrix \mathbf{A} .

After some manipulations using (7), (16) and (17), the MSE of the weighted LS channel estimation in (14) is written by

$$\varepsilon_{K+1} = \text{tr} \left[\Theta_0 + \mathbf{G}_K \Omega_K \mathbf{G}_K^\dagger - \mathbf{G}_K \Gamma_K - \Gamma_K^\dagger \mathbf{G}_K^\dagger \right], \quad (18)$$

where $\Omega_K \in \mathbb{C}^{K \times K}$ and $\Gamma_K \in \mathbb{C}^{K \times M}$ are given by

$$[\Omega_K]_{i;j} \triangleq \mathbf{x}_i^\dagger \Theta_{i-j} \mathbf{x}_j + N\sigma^2 \delta_{i-j}, \quad (19)$$

$$[\Gamma_K]_i \triangleq \mathbf{x}_i^\dagger \Theta_{i-(K+1)}. \quad (20)$$

III. OPTIMUM-WEIGHTED LEAST-SQUARES

A. Derivation of Optimum-Weighting

In this section, we optimize the weighting coefficients $\{\beta_k\}$ to minimize the MSE written in (18). We can solve the optimum values of $\{\beta_k\}$ by differentiating the MSE in terms of β_k . However, the derivative involves a highly complicated expression because β_k is included in both inside and outside of a matrix inversion of Φ_K^{-1} in \mathbf{G}_K as shown in (8). We therefore optimize the weighting matrix \mathbf{G}_K instead of optimizing the weighting coefficients $\{\beta_k\}$ in direct. Because the weighting matrix \mathbf{G}_K is constrained as $\mathbf{X}_K \mathbf{G}_K^\dagger = \mathbf{Q}_{K+1}$ mentioned in (12), we use the Lagrangian method for optimization. Notice that optimizing the weighting coefficients $\{\beta_k\}$ is equivalent to optimizing the weighting matrix \mathbf{G}_K under the condition of $\mathbf{X}_K \mathbf{G}_K^\dagger = \mathbf{Q}_{K+1}$.

Using (18) with a constraint in (12), the Lagrangian utility function is given by

$$\mathcal{L}(\mathbf{G}_K, \boldsymbol{\Upsilon}_K) \triangleq \text{tr} \left[\Theta_0 + \mathbf{G}_K \Omega_K \mathbf{G}_K^\dagger - \mathbf{G}_K \Gamma_K - \Gamma_K^\dagger \mathbf{G}_K^\dagger - 2(\mathbf{G}_K \mathbf{X}_K^\dagger - \mathbf{Q}_{K+1}^\dagger) \boldsymbol{\Upsilon}_K \right], \quad (21)$$

where $\boldsymbol{\Upsilon}_K \in \mathbb{R}^{M(D+1) \times M}$ is a Lagrangian multipliers matrix. Its derivative regarding \mathbf{G}_K^\dagger is written by

$$\frac{\partial \mathcal{L}(\mathbf{G}_K, \boldsymbol{\Upsilon}_K)}{\partial \mathbf{G}_K^\dagger} = 2 \left(\Omega_K \mathbf{G}_K^\dagger - \Gamma_K - \mathbf{X}_K^\dagger \boldsymbol{\Upsilon}_K \right). \quad (22)$$

Hence, the optimum solution is given by using the Lagrangian multipliers matrix as follows:

$$\mathbf{G}_K^\dagger = \Omega_K^{-1} \left(\Gamma_K + \mathbf{X}_K^\dagger \boldsymbol{\Upsilon}_K \right). \quad (23)$$

Substituting (23) into (12), we have

$$\mathbf{X}_K \mathbf{G}_K^\dagger = \mathbf{X}_K \Omega_K^{-1} \left(\Gamma_K + \mathbf{X}_K^\dagger \boldsymbol{\Upsilon}_K \right) = \mathbf{Q}_{K+1}. \quad (24)$$

Then, when the matrix of $(\mathbf{X}_K \Omega_K^{-1} \mathbf{X}_K^\dagger)$ is nonsingular, we obtain the optimum Lagrangian multipliers matrix as follows:

$$\boldsymbol{\Upsilon}_K = \mathbf{R}_K^{-1} \mathbf{S}_K, \quad (25)$$

where we define

$$\mathbf{R}_K \triangleq \mathbf{X}_K \Omega_K^{-1} \mathbf{X}_K^\dagger \in \mathbb{C}^{M(D+1) \times M(D+1)}, \quad (26)$$

$$\mathbf{S}_K \triangleq \mathbf{Q}_{K+1} - \mathbf{X}_K \Omega_K^{-1} \Gamma_K \in \mathbb{C}^{M(D+1) \times M}. \quad (27)$$

In consequence, by combining (23) and (25), the optimum weighting matrix for the high-order arbitrary-weighted LS channel estimation is derived as:

$$\mathbf{G}_K^\dagger = \Omega_K^{-1} \left(\Gamma_K + \mathbf{X}_K^\dagger \mathbf{R}_K^{-1} \mathbf{S}_K \right). \quad (28)$$

As described in (7), the corresponding estimation matrix is given by $\hat{\mathbf{H}}_{K+1} = \mathbf{Y}_K \mathbf{G}_K^\dagger$. The optimum weighting matrix

derived above becomes exactly equivalent to the optimum tracking weights in [21] when we assume independent PSK signals, the zero-th order LS, and the single-antenna scenario. In this view point, our derivation of the optimum weighting is a more general formulation than that in [21]; i.e. dealing with the multiple-antenna systems, any arbitrary transmission signals and the polynomial channel models.

B. Order Optimization in Optimum-Weighting

Next, we optimize the polynomial order D for the high-order OW-LS channel estimation. Through the optimization, we discover an interesting and useful result that the 0-th order OW-LS channel estimation becomes optimal. It implies that the high-order polynomial channel modeling is not advantageous any more when we can optimize the weighting coefficients.

Now, we verify the optimality of the 0-th order OW-LS in terms of minimizing the MSE. Putting the optimum weights matrix \mathbf{G}_K of (28) into the MSE of (18), the minimum MSE can be expressed as

$$\varepsilon_{K+1} = \text{tr} \left[\Theta_0 - \Gamma_K^\dagger \Omega_K^{-1} \Gamma_K + \mathbf{S}_K^\dagger \mathbf{R}_K^{-1} \mathbf{S}_K \right]. \quad (29)$$

Here, we should notice that the first two terms are independent of the estimation order D . Then, we focus on the difference between the last term $\mathbf{S}_K^\dagger \mathbf{R}_K^{-1} \mathbf{S}_K$ for the D -th order estimation and that for the $(D-1)$ -th order estimation. In this section, for convenience of the ensuing analysis, we explicitly denote the estimation order D for matrices related to the D -th order OW-LS estimation into their subscripts in parentheses; for example, $\mathbf{Q}_{k, \langle D \rangle}$. From definitions in (4) and (13), we can express $\mathbf{Q}_{k, \langle D \rangle}$ and $\mathbf{X}_{K, \langle D \rangle}$ for the D -th order estimation by using those for the $(D-1)$ -th order estimation as follows:

$$\mathbf{Q}_{k, \langle D \rangle} = \begin{bmatrix} \mathbf{Q}_{k, \langle D-1 \rangle} \\ \mathbf{Q}_{k, \langle D \rangle} \end{bmatrix}, \quad \mathbf{X}_{K, \langle D \rangle} = \begin{bmatrix} \mathbf{X}_{K, \langle D-1 \rangle} \\ \mathbf{X}_{K, \langle D \rangle} \end{bmatrix}, \quad (30)$$

where we denote $\mathbf{Q}_{k, \langle D \rangle} \triangleq k^D \mathbf{I}_M \in \mathbb{R}^{M \times M}$ and $\mathbf{X}_{K, \langle D \rangle} \triangleq [1^D \mathbf{x}_1, 2^D \mathbf{x}_2, \dots, K^D \mathbf{x}_K] \in \mathbb{C}^{M \times K}$. They lead to the following description about $\mathbf{S}_{K, \langle D \rangle}$ in (27):

$$\mathbf{S}_{K, \langle D \rangle} = \mathbf{Q}_{K+1, \langle D \rangle} - \mathbf{X}_{K, \langle D \rangle} \Omega_K^{-1} \Gamma_K = \begin{bmatrix} \mathbf{S}_{K, \langle D-1 \rangle} \\ \mathbf{S}_{K, \langle D \rangle} \end{bmatrix}, \quad (31)$$

where $\mathbf{S}_{K, \langle D \rangle} \triangleq \mathbf{Q}_{K+1, \langle D \rangle} - \mathbf{X}_{K, \langle D \rangle} \Omega_K^{-1} \Gamma_K \in \mathbb{C}^{M \times K}$. Likewise, using (30), $\mathbf{R}_{K, \langle D \rangle}$ can be expressed by

$$\begin{aligned} \mathbf{R}_{K, \langle D \rangle} &= \mathbf{X}_{K, \langle D \rangle} \Omega_K^{-1} \mathbf{X}_{K, \langle D \rangle}^\dagger \\ &= \begin{bmatrix} \mathbf{R}_{K, \langle D-1 \rangle} & \mathbf{T}_{K, \langle D \rangle} \\ \mathbf{T}_{K, \langle D \rangle}^\dagger & \mathbf{R}_{K, \langle D \rangle} \end{bmatrix}, \end{aligned} \quad (32)$$

where

$$\mathbf{T}_{K, \langle D \rangle} \triangleq \mathbf{X}_{K, \langle D-1 \rangle} \Omega_K^{-1} \mathbf{X}_{K, \langle D \rangle}^\dagger \in \mathbb{C}^{MD \times M}, \quad (33)$$

$$\mathbf{R}_{K, \langle D \rangle} \triangleq \mathbf{X}_{K, \langle D \rangle} \Omega_K^{-1} \mathbf{X}_{K, \langle D \rangle}^\dagger \in \mathbb{C}^{M \times M}. \quad (34)$$

The inversion matrix of $\mathbf{R}_{K,\langle D \rangle}$ in (32) can be written by [22]

$$\mathbf{R}_{K,\langle D \rangle}^{-1} = \begin{bmatrix} \mathbf{R}_{K,\langle D-1 \rangle}^{-1} + \mathbf{P}_{K,\langle D \rangle} \mathbf{\Pi}_{K,\langle D \rangle} \mathbf{P}_{K,\langle D \rangle}^\dagger & -\mathbf{P}_{K,\langle D \rangle} \mathbf{\Pi}_{K,\langle D \rangle} \\ -\mathbf{\Pi}_{K,\langle D \rangle} \mathbf{P}_{K,\langle D \rangle}^\dagger & \mathbf{\Pi}_{K,\langle D \rangle} \end{bmatrix}, \quad (35)$$

where

$$\mathbf{P}_{K,\langle D \rangle} \triangleq \mathbf{R}_{K,\langle D-1 \rangle}^{-1} \mathbf{T}_{K,\langle D \rangle} \in \mathbb{C}^{M \times M}, \quad (36)$$

$$\mathbf{\Pi}_{K,\langle D \rangle} \triangleq \left(\mathbf{R}_{K,\langle D \rangle} - \mathbf{T}_{K,\langle D \rangle}^\dagger \mathbf{P}_{K,\langle D \rangle} \right)^{-1} \in \mathbb{C}^{M \times M}. \quad (37)$$

Here, we have the relationships between $\mathbf{S}_{K,\langle D \rangle}$ and $\mathbf{S}_{K,\langle D-1 \rangle}$ in (31), and between $\mathbf{R}_{K,\langle D \rangle}^{-1}$ and $\mathbf{R}_{K,\langle D-1 \rangle}^{-1}$ in (35). Combining them yields the relation between the last term $\mathbf{S}_K^\dagger \mathbf{R}_K^{-1} \mathbf{S}_K$ of the MSE in (29) for the D -th order and that for the $(D-1)$ -th order as follows:

$$\begin{aligned} \mathbf{S}_{K,\langle D \rangle}^\dagger \mathbf{R}_{K,\langle D \rangle}^{-1} \mathbf{S}_{K,\langle D \rangle} &= \\ \mathbf{S}_{K,\langle D-1 \rangle}^\dagger \mathbf{R}_{K,\langle D-1 \rangle}^{-1} \mathbf{S}_{K,\langle D-1 \rangle} &+ \mathbf{P}_{K,\langle D \rangle} \mathbf{\Pi}_{K,\langle D \rangle} \mathbf{P}_{K,\langle D \rangle}^\dagger, \end{aligned} \quad (38)$$

where

$$\mathbf{P}_{K,\langle D \rangle} \triangleq \mathbf{S}_{K,\langle D-1 \rangle}^\dagger \mathbf{P}_{K,\langle D \rangle} - \mathbf{S}_{K,\langle D \rangle}^\dagger \in \mathbb{C}^{M \times M}. \quad (39)$$

Consequently, observing (29) and (38), we derive the following relation between the MSE of the D -th order estimation and that of the $(D-1)$ -th order estimation:

$$\begin{aligned} \varepsilon_{K+1,\langle D \rangle} &= \varepsilon_{K+1,\langle D-1 \rangle} + \text{tr} \left[\mathbf{P}_{K,\langle D \rangle} \mathbf{\Pi}_{K,\langle D \rangle} \mathbf{P}_{K,\langle D \rangle}^\dagger \right] \\ &= \varepsilon_{K+1,\langle D-1 \rangle} + \left\| \mathbf{P}_{K,\langle D \rangle} \mathbf{\Pi}_{K,\langle D \rangle}^{1/2} \right\|^2. \end{aligned} \quad (40)$$

Here, $\mathbf{\Pi}_{K,\langle D \rangle}^{1/2}$ denotes a matrix square root such that $\mathbf{\Pi}_{K,\langle D \rangle}^{1/2} (\mathbf{\Pi}_{K,\langle D \rangle}^{1/2})^\dagger = \mathbf{\Pi}_{K,\langle D \rangle}$. In a straightforward manner, the above relation gives $\varepsilon_{K+1,\langle 0 \rangle} \leq \varepsilon_{K+1,\langle 1 \rangle} \leq \dots \leq \varepsilon_{K+1,\langle D \rangle}$; resulting into the fact that the 0-th order OW-LS has the minimum MSE. Therefore, it is summarized in this section that the higher-order LS estimation is not effective for improving the MSE performance when we use the optimum weighting. Since the higher-order LS offers a better estimation for several channel conditions in both the rectangular weighting and the exponential weighting as shown in [19, 20], the proven fact in this section (the zero-th order LS is always optimal for the optimum weighting) may be interesting. In addition, the fact contributes to a practical use because the higher-order LS occurs higher complexity inevitably. Hereafter, we focus on the 0-th order LS channel estimation for the OW-LS channel estimation (each matrices $\{\mathbf{Q}_k\}$ just reduces into an identity matrix \mathbf{I}_M).

C. Optimum-Weighted RLS Algorithm

Since the derived optimum weighting in (28) depends on the transmitted signals and training length K , we should compute it at every symbol for decision-directed channel tracking in an extremely fast fading channel. (It should be mentioned that, in a pilot sequence period, the weighting matrix \mathbf{G}_{K_p} can be computed and memorized in advance at the receiver.) In this section, we formulate a recursive scheme like the RLS

algorithm in order to reduce the computational complexity for channel tracking; we call it as the OW-RLS algorithm. As observed in (28), there exist two matrix inversions for $\mathbf{\Omega}_K^{-1}$ and \mathbf{R}_K^{-1} . Since the matrix inversion requires a cubic complexity order in general [22], these computational complexities become $\mathcal{O}[K^3]$ and $\mathcal{O}[M^3]$, respectively. In particular, the computational burden for the matrix inversion $\mathbf{\Omega}_K^{-1}$ is a dominant factor because we often have $K \gg M$. Thus, we begin with developing an efficient recursion technique for computing $\mathbf{\Omega}_K^{-1}$.

From the definitions of $\mathbf{\Omega}_K$ in (19) and $\mathbf{\Gamma}_K$ in (20), we can write

$$\mathbf{\Omega}_K = \begin{bmatrix} \mathbf{\Omega}_{K-1} & \boldsymbol{\omega}_K \\ \boldsymbol{\omega}_K^\dagger & \varpi_K \end{bmatrix}, \quad (41)$$

where

$$\boldsymbol{\omega}_K \triangleq \mathbf{\Gamma}_{K-1} \mathbf{x}_K \in \mathbb{C}^{(K-1) \times 1}, \quad (42)$$

$$\varpi_K \triangleq \mathbf{x}_K^\dagger \boldsymbol{\Theta}_0 \mathbf{x}_K + N\sigma^2 \in \mathbb{R}. \quad (43)$$

Then, we can recursively compute the inversion in a square complexity order $\mathcal{O}[K^2]$ as follows [22]:

$$\mathbf{\Omega}_K^{-1} = \begin{bmatrix} \mathbf{\Omega}_{K-1}^{-1} + \nu_K \boldsymbol{\psi}_K \boldsymbol{\psi}_K^\dagger & -\nu_K \boldsymbol{\psi}_K \\ -\nu_K \boldsymbol{\psi}_K^\dagger & \nu_K \end{bmatrix}, \quad (44)$$

where

$$\boldsymbol{\psi}_K \triangleq \mathbf{\Omega}_{K-1}^{-1} \boldsymbol{\omega}_K \in \mathbb{C}^{(K-1) \times 1}, \quad (45)$$

$$\nu_K \triangleq \left(\varpi_K - \boldsymbol{\omega}_K^\dagger \boldsymbol{\psi}_K \right)^{-1} \in \mathbb{R}. \quad (46)$$

The above recursion requires the computational complexity of $\mathcal{O}[K^2]$ to get both $\mathbf{\Omega}_K^{-1}$ by (44) and $\boldsymbol{\psi}_K$ by (45). As discussed later, we can exclude the explicit use of the former ($\mathbf{\Omega}_K^{-1}$) from the OW-RLS algorithm by explicitly using the latter ($\boldsymbol{\psi}_K$) instead. Although $\mathbf{\Omega}_K^{-1}$ is needed for $\boldsymbol{\psi}_K$ as in (45), we can avoid the computation of (44) by making use of the modified Cholesky factorization [22]. We suppose that the matrix $\mathbf{\Omega}_K^{-1}$ is factorized as follows:

$$\mathbf{\Omega}_K^{-1} = \mathbf{L}_K \mathbf{\Lambda}_K \mathbf{L}_K^\dagger, \quad (47)$$

where $\mathbf{L}_K \in \mathbb{C}^{K \times K}$ is a lower triangular matrix whose diagonal entries are all ones, and $\mathbf{\Lambda}_K \in \mathbb{R}^{K \times K}$ is a diagonal matrix. Using (44), we obtain the relationship between \mathbf{L}_K and \mathbf{L}_{K-1} and between $\mathbf{\Lambda}_K$ and $\mathbf{\Lambda}_{K-1}$ as follows:

$$\begin{aligned} \mathbf{\Omega}_K^{-1} &= \begin{bmatrix} \mathbf{L}_{K-1} \mathbf{\Lambda}_{K-1} \mathbf{L}_{K-1}^\dagger + \nu_K \boldsymbol{\psi}_K \boldsymbol{\psi}_K^\dagger & -\nu_K \boldsymbol{\psi}_K \\ -\nu_K \boldsymbol{\psi}_K^\dagger & \nu_K \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \mathbf{L}_{K-1} & \boldsymbol{\psi}_K \\ \mathbf{0}_{K-1}^\top & 1 \end{bmatrix}}_{\mathbf{L}_K} \underbrace{\begin{bmatrix} \mathbf{\Lambda}_{K-1} & \mathbf{0}_{K-1} \\ \mathbf{0}_{K-1}^\top & \nu_K \end{bmatrix}}_{\mathbf{\Lambda}_K} \underbrace{\begin{bmatrix} \mathbf{L}_{K-1}^\dagger & \mathbf{0}_{K-1} \\ \boldsymbol{\psi}_K^\dagger & 1 \end{bmatrix}}_{\mathbf{L}_K^\dagger}. \end{aligned} \quad (48)$$

As a result, the Cholesky factorized matrices \mathbf{L}_K and $\mathbf{\Lambda}_K$ do not require any computations for the recursive process. Using these matrices, the vector $\boldsymbol{\psi}_K$ in (45) is rewritten as

$$\boldsymbol{\psi}_K = \mathbf{L}_{K-1} \mathbf{\Lambda}_{K-1} \mathbf{L}_{K-1}^\dagger \boldsymbol{\omega}_K. \quad (49)$$

Notice that the computational burden in (49) is exactly identical to that in (45), while the recursive calculation of $\mathbf{\Omega}_K^{-1}$ by (44) is omitted.

Next, we reduce the computational complexity for the matrix inversion \mathbf{R}_K^{-1} to a square order $\mathcal{O}[M^2]$ from a cubic order $\mathcal{O}[M^3]$ by using a recursive way. Let \mathbf{U}_K and \mathbf{V}_K be defined as follows:

$$\mathbf{U}_K \triangleq \mathbf{X}_K \mathbf{\Omega}_K^{-1} \in \mathbb{C}^{M \times K}, \quad (50)$$

$$\mathbf{V}_K \triangleq \mathbf{Y}_K \mathbf{\Omega}_K^{-1} \in \mathbb{C}^{N \times K}. \quad (51)$$

They are recursively obtained by (44):

$$\begin{aligned} \mathbf{U}_K &= [\mathbf{X}_{K-1} \quad \mathbf{x}_K] \begin{bmatrix} \mathbf{\Omega}_{K-1}^{-1} + \nu_K \boldsymbol{\psi}_K \boldsymbol{\psi}_K^\dagger & -\nu_K \boldsymbol{\psi}_K \\ -\nu_K \boldsymbol{\psi}_K^\dagger & \nu_K \end{bmatrix} \\ &= [\mathbf{U}_{K-1} - \nu_K \boldsymbol{\xi}_K \boldsymbol{\psi}_K^\dagger \quad \nu_K \boldsymbol{\xi}_K], \end{aligned} \quad (52)$$

$$\begin{aligned} \mathbf{V}_K &= [\mathbf{Y}_{K-1} \quad \mathbf{y}_K] \begin{bmatrix} \mathbf{\Omega}_{K-1}^{-1} + \nu_K \boldsymbol{\psi}_K \boldsymbol{\psi}_K^\dagger & -\nu_K \boldsymbol{\psi}_K \\ -\nu_K \boldsymbol{\psi}_K^\dagger & \nu_K \end{bmatrix} \\ &= [\mathbf{V}_{K-1} - \nu_K \boldsymbol{\varphi}_K \boldsymbol{\psi}_K^\dagger \quad \nu_K \boldsymbol{\varphi}_K], \end{aligned} \quad (53)$$

where we defined

$$\boldsymbol{\xi}_K \triangleq \mathbf{x}_K - \mathbf{X}_{K-1} \boldsymbol{\psi}_K = \mathbf{S}_{K-1} \mathbf{x}_K \in \mathbb{C}^{M \times 1}, \quad (54)$$

$$\boldsymbol{\varphi}_K \triangleq \mathbf{y}_K - \mathbf{Y}_{K-1} \boldsymbol{\psi}_K \in \mathbb{C}^{N \times 1}. \quad (55)$$

Using (50), (52) and (54), we can write the matrix $\mathbf{R}_K = \mathbf{X}_K \mathbf{\Omega}_K^{-1} \mathbf{X}_K^\dagger$ in a recursive manner as follows:

$$\begin{aligned} \mathbf{R}_K &= \mathbf{U}_K \mathbf{X}_K^\dagger \\ &= [\mathbf{U}_{K-1} - \nu_K \boldsymbol{\xi}_K \boldsymbol{\psi}_K^\dagger \quad \nu_K \boldsymbol{\xi}_K] \begin{bmatrix} \mathbf{X}_{K-1}^\dagger \\ \mathbf{x}_K^\dagger \end{bmatrix} \\ &= \mathbf{R}_{K-1} + \nu_K \boldsymbol{\xi}_K \boldsymbol{\xi}_K^\dagger. \end{aligned} \quad (56)$$

Accordingly, its inversion can be recursively computed with a square complexity order $\mathcal{O}[M^2]$ by the Sherman-Morrison formula [22]:

$$\mathbf{R}_K^{-1} = \mathbf{R}_{K-1}^{-1} - \mu_K \boldsymbol{\eta}_K \boldsymbol{\eta}_K^\dagger, \quad (57)$$

where

$$\boldsymbol{\eta}_K \triangleq \mathbf{R}_{K-1}^{-1} \boldsymbol{\xi}_K \in \mathbb{C}^{M \times 1}, \quad (58)$$

$$\mu_K \triangleq \left(1/\nu_K + \boldsymbol{\xi}_K^\dagger \boldsymbol{\eta}_K\right)^{-1} \in \mathbb{R}. \quad (59)$$

Observed in (7), (28) and (51), the estimated matrix by the OW-LS is written as $\hat{\mathbf{H}}_{K+1} = \mathbf{V}_K \boldsymbol{\Gamma}_K + \boldsymbol{\Sigma}_K \mathbf{S}_K$, where we defined $\boldsymbol{\Sigma}_K \triangleq \mathbf{V}_K \mathbf{X}_K^\dagger \mathbf{R}_K^{-1} \in \mathbb{C}^{N \times M}$. We showed that some parts of $\hat{\mathbf{H}}_{K+1}$ can be recursively generated; \mathbf{V}_K by (53), \mathbf{R}_K^{-1} included in $\boldsymbol{\Sigma}_K$ by (57), and \mathbf{U}_K included in \mathbf{S}_K by (52). Finally, we show the recursive generation of $\boldsymbol{\Sigma}_K$. From the definition, $\boldsymbol{\Sigma}$ is written by

$$\begin{aligned} \boldsymbol{\Sigma}_K &= \mathbf{V}_K \mathbf{X}_K^\dagger \mathbf{R}_K^{-1} \\ &= [\mathbf{V}_{K-1} - \nu_K \boldsymbol{\varphi}_K \boldsymbol{\psi}_K^\dagger \quad \nu_K \boldsymbol{\varphi}_K] \begin{bmatrix} \mathbf{X}_{K-1}^\dagger \\ \mathbf{x}_K^\dagger \end{bmatrix} \mathbf{R}_K^{-1} \\ &= \left(\mathbf{V}_{K-1} \mathbf{X}_{K-1}^\dagger + \nu_K \boldsymbol{\varphi}_K \boldsymbol{\xi}_K^\dagger\right) \left(\mathbf{R}_{K-1}^{-1} - \mu_K \boldsymbol{\eta}_K \boldsymbol{\eta}_K^\dagger\right) \\ &= \boldsymbol{\Sigma}_{K-1} + \mu_K (\boldsymbol{\varphi}_K - \boldsymbol{\Sigma}_{K-1} \boldsymbol{\xi}_K) \boldsymbol{\eta}_K^\dagger \\ &= \boldsymbol{\Sigma}_{K-1} + \mu_K \mathbf{e}_K \boldsymbol{\eta}_K^\dagger. \end{aligned} \quad (60)$$

Here, we used $\boldsymbol{\xi}_K^\dagger \boldsymbol{\eta}_K = \mu_K^{-1} - \nu_K^{-1}$ from (59). Making use of these recursions, we propose a recursive channel tracking scheme called as the OW-RLS algorithm; whose overall procedure is described below:

- 1: set empty: $\boldsymbol{\Gamma}_0, \mathbf{L}_0, \mathbf{A}_0, \mathbf{U}_0, \mathbf{Y}_0, \mathbf{V}_0, \hat{\mathbf{H}}_1, \boldsymbol{\Sigma}_0$
- 2: set diagonal: $\mathbf{S}_0 = \mathbf{I}_M, \mathbf{R}_0^{-1} = \rho \mathbf{I}_M$
- 3: estimate: $\boldsymbol{\Theta}_i$ for $0 \leq i \leq K_p + K_i$
- 4: **for** $K = 1$ to $K_p + K_i - 1$ **do**
- 5: $\mathbf{e}_K = \mathbf{y}_K - \hat{\mathbf{H}}_K \mathbf{x}_K$
- 6: $\boldsymbol{\omega}_K = \boldsymbol{\Gamma}_{K-1} \mathbf{x}_K$
- 7: $\boldsymbol{\psi}_K = \mathbf{L}_{K-1} \mathbf{A}_{K-1} \mathbf{L}_{K-1}^\dagger \boldsymbol{\omega}_K$
- 8: $\boldsymbol{\xi}_K = \mathbf{S}_{K-1} \mathbf{x}_K$
- 9: $\boldsymbol{\eta}_K = \mathbf{R}_{K-1}^{-1} \boldsymbol{\xi}_K$
- 10: $\boldsymbol{\varphi}_K = \mathbf{e}_K + \boldsymbol{\Sigma}_{K-1} \boldsymbol{\xi}_K$
- 11: $\nu_K = 1/(N\sigma^2 + \mathbf{x}_K^\dagger \boldsymbol{\Theta}_0 \mathbf{x}_K - \boldsymbol{\omega}_K^\dagger \boldsymbol{\psi}_K)$
- 12: $\mu_K = 1/(1/\nu_K + \boldsymbol{\xi}_K^\dagger \boldsymbol{\eta}_K)$
- 13: $\mathbf{L}_K = \begin{bmatrix} \mathbf{L}_{K-1} & \boldsymbol{\psi}_K \\ \mathbf{0}_{K-1}^\top & 1 \end{bmatrix}$
- 14: $\mathbf{A}_K = \begin{bmatrix} \mathbf{A}_{K-1} & \mathbf{0}_{K-1} \\ \mathbf{0}_{K-1}^\top & \nu_K \end{bmatrix}$
- 15: $\mathbf{U}_K = [\mathbf{U}_{K-1} - \nu_K \boldsymbol{\xi}_K \boldsymbol{\psi}_K^\dagger \quad \nu_K \boldsymbol{\xi}_K]$
- 16: $\mathbf{V}_K = [\mathbf{V}_{K-1} - \nu_K \boldsymbol{\varphi}_K \boldsymbol{\psi}_K^\dagger \quad \nu_K \boldsymbol{\varphi}_K]$
- 17: $\mathbf{R}_K^{-1} = \mathbf{R}_{K-1}^{-1} - \mu_K \boldsymbol{\eta}_K \boldsymbol{\eta}_K^\dagger$
- 18: $\boldsymbol{\Sigma}_K = \boldsymbol{\Sigma}_{K-1} + \mu_K \mathbf{e}_K \boldsymbol{\eta}_K^\dagger$
- 19: generate $\boldsymbol{\Gamma}_K$
- 20: $\mathbf{S}_K = \mathbf{I}_M - \mathbf{U}_K \boldsymbol{\Gamma}_K$
- 21: $\hat{\mathbf{H}}_{K+1} = \mathbf{V}_K \boldsymbol{\Gamma}_K + \boldsymbol{\Sigma}_K \mathbf{S}_K$
- 22: **end for**

Although the proposed algorithm here has still high complexity in comparison with the classical EW-RLS algorithm, the recursive algorithm can decrease the complexity to the square order from the cubic order of the direct OW-LS and, moreover, the algorithm for the channel tracking has originality with a quite difference from the method for a single-antenna case in [21], in which any recursive formulation has not been used for channel tracking, assuming the optimum weights are computed in advance.

IV. PERFORMANCE EVALUATIONS

In this section, we demonstrate the advantage of the proposed OW-RLS channel estimation scheme in MSE and bit-error-rate (BER) performance. Here, we use two MIMO detection algorithms: the maximum-likelihood detection (MLD) [24] as an optimum receiving algorithm, and the vertical Bell laboratories layered space-time architecture (V-BLAST) which is based on the minimum MSE criterion with an optimum ordering [25]. We consider a 4×4 MIMO system and a spatially correlated fading channel that is modeled by the Kronecker representation $\mathbf{H}_k = \mathbf{R}_{\text{rx}}^{1/2} \check{\mathbf{H}}_k \mathbf{R}_{\text{tx}}^{\text{T}/2}$ [26–29]. Both of the transmitter correlation \mathbf{R}_{tx} and the receiver correlation \mathbf{R}_{rx} have a correlation factor of 0.2 at their non-diagonal entries, and $\check{\mathbf{H}}_k$ generates an uncorrelated Rayleigh fading by Jakes model with the maximum Doppler frequency of f_D . We assume the transmission frame consists of a pilot sequence of $K_p = 32$ symbols and an information sequence of $K_i = 128$ symbols using 16QAM modulation schemes. Pilot symbols are spatially orthogonal Walsh sequences in BPSK modulation.

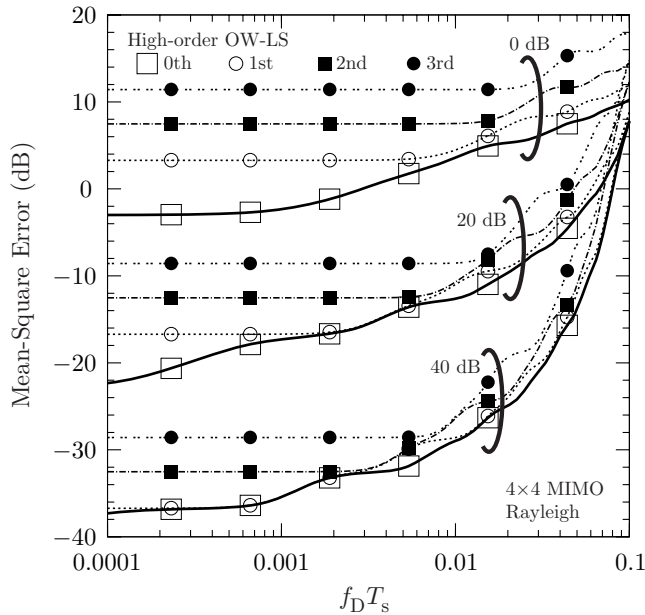


Fig. 1. MSE performance as a function of $f_D T_s$ for the D -th order OW-LS (4×4 MIMO system, average SNR = 0 dB, 20 dB and 40 dB, $K_p = 32$, $D = 0, 1, 2, 3$). Plotted curves represent theoretical results, and markers are simulation results.

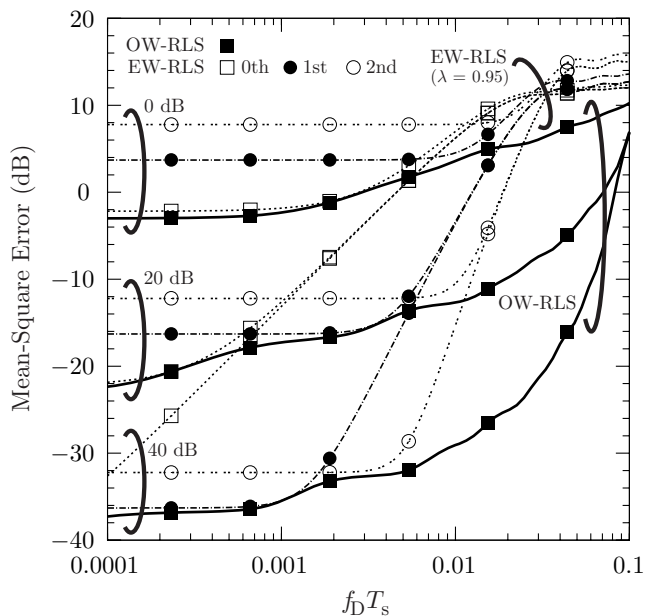


Fig. 2. MSE performance as a function of $f_D T_s$ for OW-RLS and the D -th order EW-RLS with a forgetting factor of 0.95 (4×4 MIMO system, average SNR = 0 dB, 20 dB and 40 dB, $K_p = 32$). Plotted curves represent theoretical results, and markers are simulation results.

A. MSE Performance

First of all, we show a simulation result to back up the proved theory that the 0-th order OW-LS is optimum among the D -th order OW-LS. Fig. 1 shows the MSE performance ε_{K_p+1} of the D -th order OW-LS as a function of $f_D T_s$ for an average SNR of 0 dB, 20 dB and 40 dB. Plotted curves represent the theoretical results given in (18), and plotted markers are the simulation results. One can see that the simulation results well match the theoretical curves. It is

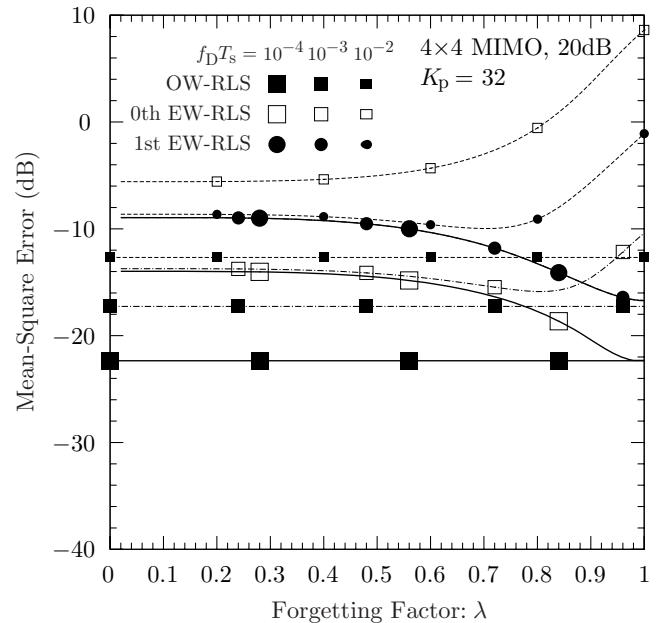


Fig. 3. MSE performance as a function of a forgetting factor λ for average SNR of 20 dB (4×4 MIMO system, $K_p = 32$, $f_D T_s = 0.0001$ and 0.01). Plotted curves represent theoretical results, and markers are simulation results.

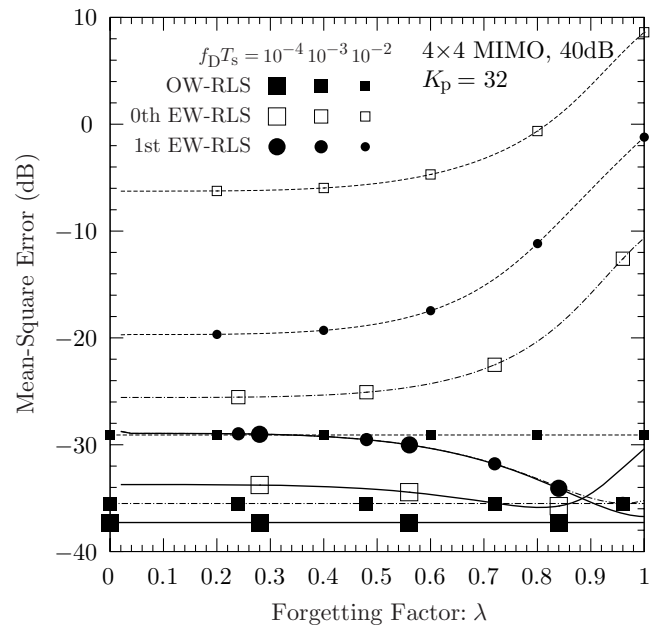


Fig. 4. MSE performance as a function of a forgetting factor λ for average SNR of 40 dB (4×4 MIMO system, $K_p = 32$, $f_D T_s = 0.0001$ and 0.01). Plotted curves represent theoretical results, and markers are simulation results.

inevitable that the MSE performance gets worse as the fading speed $f_D T_s$ increases. As shown in this figure, the lower-order OW-LS offers the better MSE performance, and the 0-th order OW-LS is the best for any SNR and $f_D T_s$. It can be confirmed that the high-order polynomial channel modeling [19, 20] is not useful when we can optimize the weighting coefficients. Notice that, this figure shows the lower bound of the D -th order LS channel estimation: in other words, any arbitrary-weighted LS channel estimation with the D -th order polynomial model cannot outperform these plotted curves.

Next, we show the comparison between the OW-RLS and

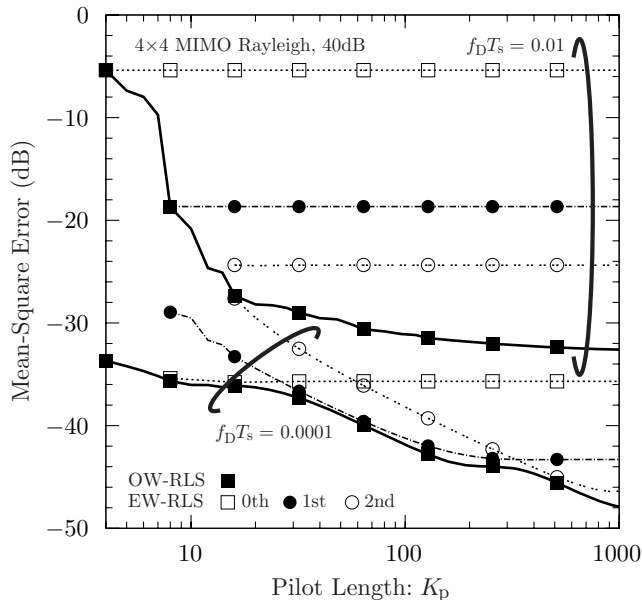


Fig. 5. MSE performance as a function of pilot sequence length K_p for OW-RLS and the D -th order EW-RLS with optimized forgetting factor (4×4 MIMO system, average SNR = 40 dB, $f_D T_s = 0.0001$ and 0.01). Plotted curves represent theoretical results, and markers are simulation results.

the D -th order EW-RLS channel estimation. In Fig. 2, we plot the MSE performance ε_{K_p+1} versus a fading speed $f_D T_s$ for the D -th order EW-RLS with a forgetting factor of $\lambda = 0.95$, along with for the (0-th order) OW-RLS. Observed in this figure, the lower-order EW-RLS channel estimation is severely degraded in fast fading and at high SNR region. And then, the performance degradation can be overcome by the higher-order polynomial channel modeling in the exponential weighting. Note that, although the higher-order EW-RLS is advantageous for rapid fading, it has an adverse effect at low SNR and in slow fading channels. This is because the higher-order estimation requires a longer pilot sequence for convergence. Therefore, we should properly choose the polynomial order for the high-order EW-RLS algorithm, depending on the channel conditions and the pilot length. As is obvious, the OW-RLS always outperforms the D -th order EW-RLS; the performance gain is significant especially in rapid fading channels.

The MSE performance of the EW-RLS channel estimation may be improved by optimizing the forgetting factor for fast fading channels. We show the MSE performance of the EW-RLS as a function of a forgetting factor λ in Figs. 3 and 4 for an average SNR of 20 dB and 40 dB, respectively. We plot the MSE performance for $f_D T_s = 0.0001, 0.001$ and 0.01. For reference, the MSE performance of the OW-RLS is presented in these figures. (Note that the performance is independent of the forgetting factor because the OW-RLS does not use the forgetting factor.) Here, it should be mentioned that the MSE performance greatly depends on the forgetting factor, and also the optimum value of the forgetting factor highly depends on the channel conditions such as SNR and $f_D T_s$. Therefore, we should properly adjust the forgetting factor according to the channel environment in an experimental way since the optimum forgetting factor of the high-order EW-RLS has not been theoretically derived for any arbitrary channel statistics.

In slow fading such as $f_D T_s = 0.0001$, since the estimation error due to the additive noise is dominant especially at low SNR, a larger value of the forgetting factor offers a better MSE performance. At low SNR of 20 dB for such a slow fading, the 0-th order EW-RLS is superior to the 1-st order one because the noise error is a dominant factor, whereas the 0-th order one is inferior to the 1-st order one at high SNR of 40 dB because the lag error can be larger than the noise error. It confirms that the higher-order EW-RLS is useful to suppress the lag error caused by the channel variation compared to the lower-order one, while it works negatively when the lag error is outweighed by the noise error. In fast fading such as $f_D T_s = 0.01$, a relatively small value of the forgetting factor is required because the lag error can surpass the noise error. For such a fast fading, the higher-order EW-RLS contributes to improving the MSE performance. The OW-RLS offers further performance improvement especially in rapid fading channels because it adaptively adjusts the weighting coefficient in favor of minimizing the total estimation error of the noise error and the lag error by exploiting the channel statistics.

As shown in Figs. 2 through 4, the higher-order EW-RLS cannot always have a better performance than the lower-order EW-RLS because it requires a longer pilot sequence for convergence, owing to the larger number of parameters to be estimated. Now, we show the influence of the pilot length in Fig. 5 for an average SNR of 40 dB. Because the inverse matrix of Φ_{K_p} in (9) does not exist for $K_p < M(D+1)$, we plot the MSE performance from $K_p \geq M(D+1)$ for the D -th order estimation. The forgetting factor is manually optimized for every plotted point. In rapid fading for $f_D T_s = 0.01$, a substantial advantage of the higher-order EW-RLS can be seen because the lag error is a dominant factor. Meanwhile, in slow fading, because the larger number of estimation parameters may incur an enhancement of the noise error, the MSE performance of the higher-order EW-RLS gets worse for a shorter pilot length. However, one can see that the longer pilot sequence contributes to decreasing the noise error for such a slow fading. This figure reveals that the MSE performance of the higher-order EW-RLS can approach that of the OW-RLS when we can use a longer pilot enough to converge. Nonetheless, the OW-RLS is still advantageous especially for rapid fading.

In this section, we have presented a performance evaluation at an average SNR of 0 dB, 20 dB and 40 dB just for a reference. In some practical applications, such a high SNR around 20 dB and 40 dB may not be realistic. However, as we show the BER performance in the following section, the receiver employing the MLD and V-BLAST algorithms require around 20 dB and 40 dB in SNR, respectively, to guarantee BER less than 10^{-4} for the uncoded 16QAM 4×4 MIMO systems. Hence, we may encounter with such a high SNR requirement for some applications, e.g. in which no diversity techniques or no error-correcting codes are available due to the system cost.

B. BER Performance

Next, we evaluate the effect of the proposed OW-RLS channel estimation in terms of BER performance. Figs. 6 and 7

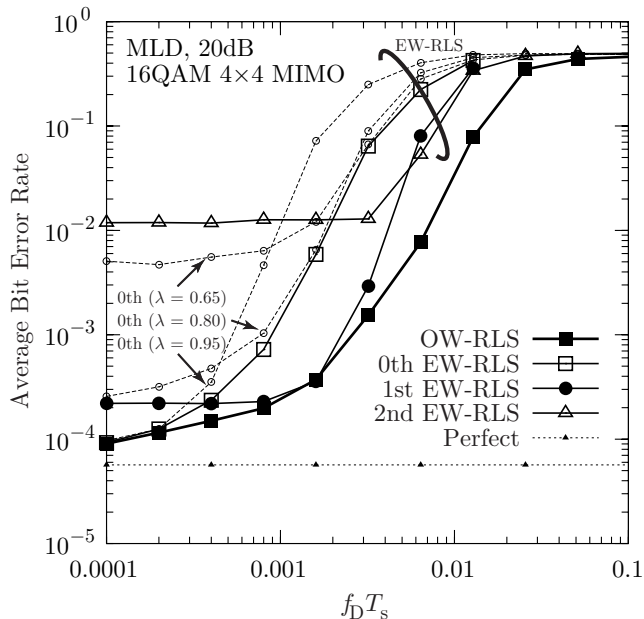


Fig. 6. Average BER performance as a function of $f_D T_s$ (MLD detection, 16QAM 4×4 MIMO system, average SNR = 20 dB).

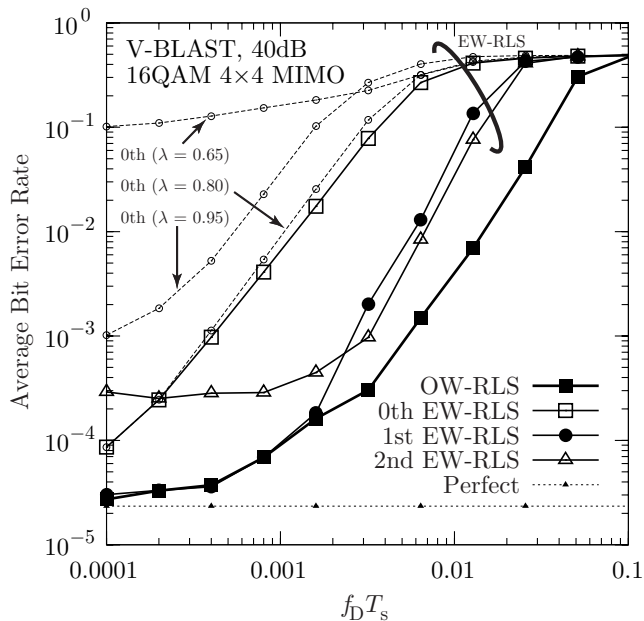


Fig. 7. Average BER performance as a function of $f_D T_s$ (V-BLAST detection, 16QAM 4×4 MIMO system, average SNR = 40 dB).

show the BER performance of MLD and V-BLAST detections as a function of $f_D T_s$ for an average SNR of 20 dB and 40 dB, respectively. For the EW-RLS estimation, we manually optimized a forgetting factor λ at every simulation points. In these figures, we also present the BER performance in the case that the channel state information is perfectly known at the receiver, for reference. (Its performance is independent of the fading speed.) Besides, we plot the BER performance of the EW-RLS when a fixed value of the forgetting factor is used ($\lambda = 0.95, 0.80,$ and 0.65). As shown in these figures, the BER performance is severely degraded in fast fading channels. Although the higher-order EW-RLS can overcome

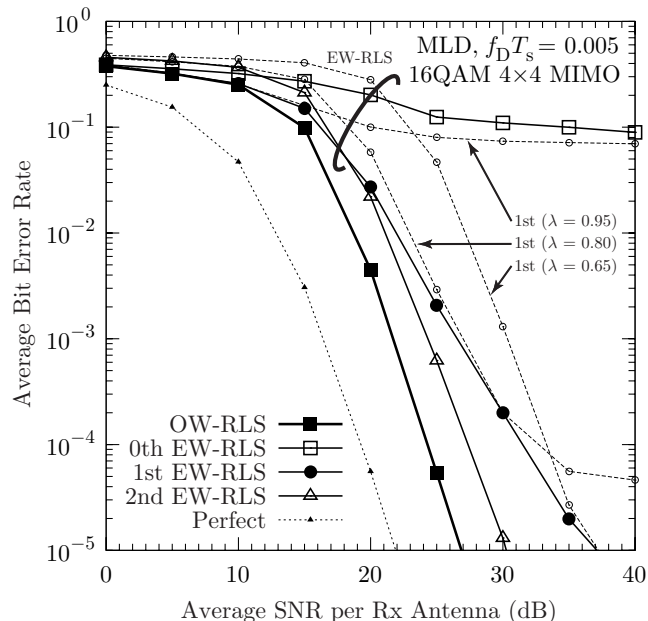


Fig. 8. Average BER performance as a function of average SNR per Rx antenna for $f_D T_s = 0.005$ (MLD detection, 16QAM 4×4 MIMO system).

such a performance degradation, it is inferior to the 0-th order one in slow fading channels and at low SNR. The trend of the BER performance basically matches that of the MSE performance shown in Fig. 2. Comparing the BER curves with a fixed forgetting factor to those with an optimized factor, it is confirmed that we need a careful adjustment of the forgetting factor in accord with the channel environments. We can see that the proposed OW-RLS has much better tolerance for fast fading MIMO channels resulting from high mobility.

Fig. 8 shows the BER performance versus average SNR for fast fading ($f_D T_s = 0.005$), in which the MLD receiver is employed. In the D -th order EW-RLS, the forgetting factor is well optimized at every point. It should be noticed that the 0-th order EW-RLS has a significantly poor performance in such a rapid fading, in spite of the forgetting factor is optimized. The higher-order EW-RLS can improve the BER performance considerably. However, observing the BER performance of the 1-st order EW-RLS with a fixed forgetting factor ($\lambda = 0.95, 0.80,$ and 0.65), it can be confirmed that the BER performance is severely degraded unless the forgetting factor is well adjusted according to an SNR. Although the proposed OW-RLS has an approximately 5 dB loss from an ideal performance with a perfect channel knowledge, it yields the best BER performance over all the performance with a channel estimation; the SNR gains of 10 dB and 3 dB are obtained in comparison to the 1-st order and 2-nd order EW-RLS, respectively, at a BER of 10^{-5} .

C. Impact of Correlation Factor and Decision/Estimation Error

Before now, we have evaluated the MSE and the BER performance for a spatial correlation factor of 0.2. In this section, we show the impact of the correlation factor to the performance. Since the OW-RLS exploits a spatial correlation information as well as the temporal correlation for the useful

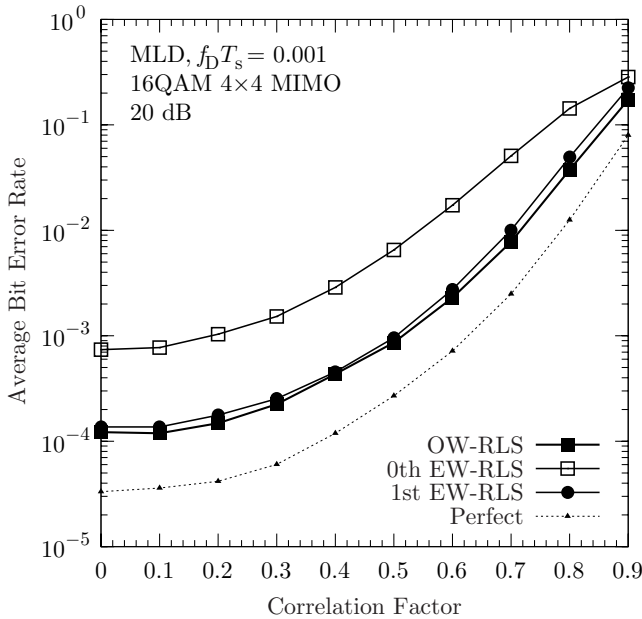


Fig. 9. Average BER performance as a function of a correlation factor for $f_D T_s = 0.001$ (MLD detection, 16QAM 4×4 MIMO system, average SNR = 20 dB).

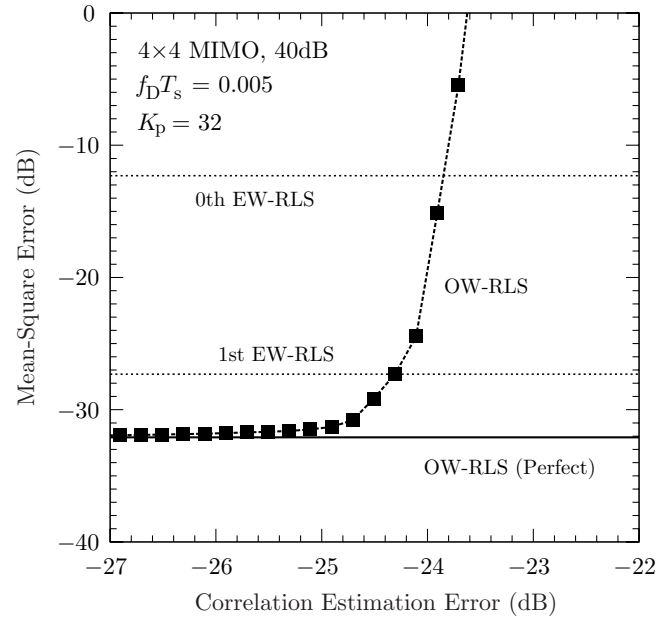


Fig. 11. MSE performance as a function of a correlation estimation error for $f_D T_s = 0.005$ (4×4 MIMO system, average SNR = 40 dB).

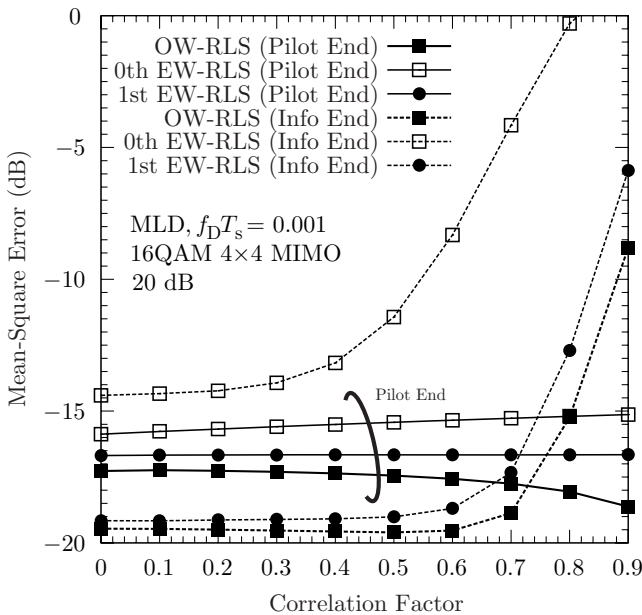


Fig. 10. MSE performance as a function of a correlation factor for $f_D T_s = 0.001$ (MLD detection, 16QAM 4×4 MIMO system, average SNR = 20 dB).

channel statistics, the MSE performance is expected to be improved for a highly correlated channel.

Fig. 9 exhibits the BER performance as a function of a correlation factor for an average SNR of 20 dB and a fading speed of $f_D T_s = 0.001$. Observing the performance with a perfect channel information in this figure, we can know that the BER performance is inevitably degraded as the spatial correlation factor increases. Although the OW-RLS has the best BER performance, any significant advantage owing to the spatial correlation cannot be confirmed.

In Fig. 10, we plot the MSE performance at both the end of the pilot sequence and the end of the information

sequence, as a function of the spatial correlation factor. It should be remarked that the OW-RLS can improve the MSE performance at the end of the pilot sequence as the correlation factor increases, whereas the EW-RLS is almost independent of the correlation factor. Therefore, it is confirmed that the OW-RLS can put the correlation factor to good use for channel estimation. However, the MSE at the end of the information sequence can be extremely degraded in a highly correlated channel because the decision error is not negligible as shown in Fig. 9. The channel tracking may improve the MSE performance when the decision error rate is sufficiently low; i.e. a BER of less than 10^{-2} for the 1-st order EW-RLS and the OW-RLS channel estimations, seen from Figs. 9 and 10.

For practical applications, the channel statistics of spatio-temporal correlations $\{\Theta_k\}$ and a noise variance σ^2 , which are required for the OW-RLS, cannot be perfectly known. Therefore, the channel estimation accuracy may be severely affected by the imperfect knowledge of the channel statistics. To evaluate the influence of the imperfection, we show the MSE performance ε_{K_p+1} of the OW-RLS in Fig. 11 as a function of a correlation estimation error for an average SNR of 40 dB and a fading speed of $f_D T_s = 0.005$. Here, we defined the correlation estimation error as $\sum_{k=1}^{K_p} \|\hat{\Theta}_k - \Theta_k\|^2$, where each $\hat{\Theta}_k$ is an estimated version of $\{\Theta_k\}$, whose entries are added by independent Gaussian noises. For comparison, this figure also exhibits the MSE performance of the OW-RLS with a perfect knowledge of the channel statistics, and the D -th order EW-RLS with an optimized forgetting factor. (Note that they are independent of the correlation estimation error.) As shown in this figure, one can see that the MSE performance is significantly degraded in the presence of a large estimation error of the spatio-temporal correlation. To obtain the performance gain against the 1-st order EW-RLS with an optimized forgetting factor, we need a correlation estimation

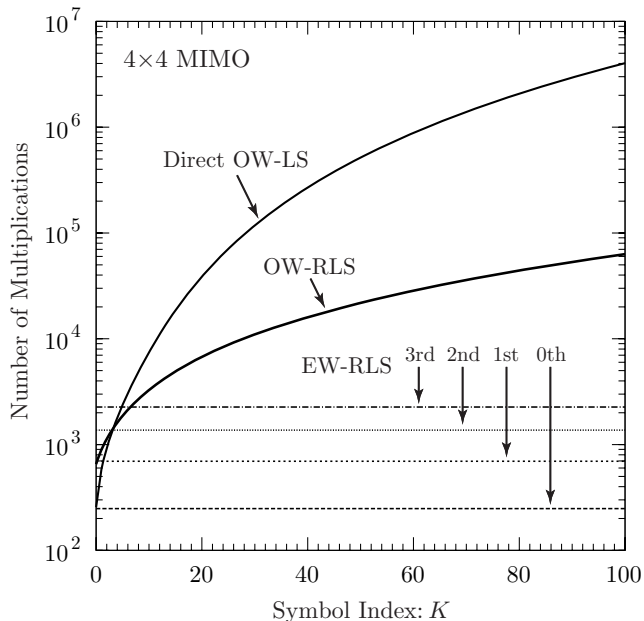


Fig. 12. Computational complexity in arithmetic multiplication versus symbol index K (4×4 MIMO system).

error less than -24.4 dB. Although its requirement may be somehow tight, an estimation error larger than the noise level by approximately 15 dB is still acceptable.

D. Computational Complexity

Table I lists the computational complexity in the required number of real-valued multiplications for the D -th order EW-RLS algorithm and the proposed OW-RLS algorithm. For reference, the computational complexity of the direct OW-LS formula derived in (28) is presented in this table, where we use the Gauss-Jordan method [22] for matrix inversion. In Table II, we show some examples for these computational complexities in a 4×4 MIMO system. For convenience in comparison, the complexity curve versus the symbol index K is plotted in Fig. 12. One can see that the proposed scheme has higher computational complexity than the EW-RLS algorithm; the computational cost of the OW-RLS algorithm roughly increases in a square order according to the training length K . It is the main drawback for the OW-RLS that the required computational complexity increases for the channel tracking process. Nevertheless, as we have evaluated the MSE and the BER performance before, the proposed scheme offers a good accuracy in channel estimation under fast fading. Furthermore, we should note that the computational complexity of the direct OW-LS without the proposed recursion technique becomes extremely high, specifically in a cubic order against K . In consequence, the proposed recursion technique is efficient to reduce the complexity of OW-LS for an accurate channel estimation.

V. OTHER DISCUSSIONS

A. Estimation of Channel Correlations

Since the proposed OW-RLS algorithm requires channel correlation matrices $\{\Theta_k\}$, we should estimate them in advance. When the channel follows a rich-scattered Rayleigh

fading, since the correlation matrices $\{\Theta_k\}$ depend only on the maximum Doppler frequency f_D , some estimation schemes of f_D , e.g. reported in [30–33], are effective to obtain the correlation matrices. Through the evaluations in the previous section, we have only assumed the classical Rayleigh fading model for a time-varying MIMO fading channel. However, the channel statistics can be complicated, depending on the surrounding scatterers, the antenna configurations, the directional characteristics of antennas, the distance of transmitter and receiver nodes, and the movement or rotation of the mobile station, as reported in [34, 35]. Moreover, when the transmit antennas are distributed to different nodes in the so-called distributed antenna or cooperative communications scenarios [36–41], the spatio-temporal correlation becomes further complicated because a channel response from each transmit antenna has an individual maximum Doppler frequency. For these reasons, we have focused on using any arbitrary channel correlations in deriving the OW-RLS algorithm for a realistic application. Whereas the method in [21] assumed that each channel coefficient has the identical Doppler spectrum: which can violate the optimality of the weights for some realistic channel environments.

One of the simplest correlation estimation schemes is as follows. Using previously estimated matrices in the subsequent transmitted frames, the correlation can be estimated as

$$\hat{\Theta}_i \simeq \frac{1}{K'} \sum_{k=1}^{K'} \hat{\mathbf{H}}_k^\dagger \hat{\mathbf{H}}_{k-i}, \quad (61)$$

where K' denotes the number of available estimated channel matrices to compute them; which should be sufficiently large for averaging. As K' increases, the estimation accuracy is improved by averaging the estimation error and the instantaneous channel correlation. Since the above-mentioned estimation scheme of channel correlation is not so sophisticated, more accurate scheme may be needed. However, some sort of a correlation estimation scheme are beyond of the main focus of this paper. The main contribution of this paper is the theoretical derivation of the exact optimum weighting for the LS channel estimation in arbitrary time-varying MIMO fading channels. As long as we can use sufficiently accurate channel statistics, a theoretically best MSE performance for channel estimation can be obtained by the proposed OW-RLS.

B. Extension to Wideband Scenarios

This paper has assumed a narrow-band transmission signal resulting into a frequency-flat fading channel. So, we should extend it to frequency-selective channels for wideband transmissions. It is natural that we use a conjunction of multi-carrier orthogonal frequency-division multiplexing (OFDM) with an MIMO system because it transforms a frequency-selective channels to multiple frequency-flat channels in subcarrier. Hence, we can use the proposed channel estimation scheme for each subcarrier in a straightforward manner. However, the OFDM signaling has some impairments: energy loss of guard intervals, requirement of coding schemes for exploiting path/frequency diversity, sensitivity of carrier synchronizations, and high peak-to-average power ratio leading to necessity of some sort of power amplifiers with a

TABLE I
COMPUTATIONAL COMPLEXITY IN MULTIPLICATIONS AT THE K -TH SYMBOL

D -th order EW-RLS	$7M^2(D+1)^2 + (6N+4)M(D+1) + 2N(N-1)$
Proposed OW-RLS	$4K^2 + (4NM + 8M^2 + 4N + 6M - 8)K + 4NM^2 + 12NM + 14M^2 - 2N - 4M + 4$
Direct OW-LS	$4K^3 + 8MK^2 + 4(3M^2 + NM)K + 4M^3$

TABLE II
EXAMPLE COMPLEXITY IN MULTIPLICATIONS FOR 4×4 MIMO SYSTEMS

D -th order EW-RLS	$112D^2 + 336D + 248$	$= 2.5 \times 10^2 (D=0), 7.0 \times 10^2 (D=1), 1.4 \times 10^3 (D=2), 2.3 \times 10^3 (D=3)$
Proposed OW-RLS	$4K^2 + 224K + 652$	$= 2.7 \times 10^3 (K=8), 1.2 \times 10^4 (K=32), 3.1 \times 10^4 (K=64)$
Direct OW-LS	$4K^3 + 32K^2 + 128K + 128$	$= 5.2 \times 10^3 (K=8), 1.7 \times 10^5 (K=32), 1.2 \times 10^6 (K=64)$

good linearity [42]. In addition, a considerable performance degradation can be occurred due to inter-carrier interference in fast fading channels.

In actual, our scheme is readily applicable to single-carrier transmissions as well as multi-carrier transmissions for wide-band scenarios. Let $\mathbf{H}_{k,l} \in \mathbb{C}^{N \times M}$ be the l -th delayed channel matrix at the k -th symbol. In frequency-selective fading channels with a $(L+1)$ -path model, one can write a received signals vector as

$$\begin{aligned} \mathbf{y}_k &= \sum_{l=0}^L \mathbf{H}_{k,l} \mathbf{x}_{k-l} + \mathbf{w}_k \\ &= \bar{\mathbf{H}}_k \bar{\mathbf{x}}_k + \mathbf{w}_k, \end{aligned} \quad (62)$$

where L is the maximum delay in symbol, and we defined as

$$\bar{\mathbf{H}}_k \triangleq [\mathbf{H}_{k,0}, \mathbf{H}_{k,1}, \dots, \mathbf{H}_{k,L}] \in \mathbb{C}^{N \times M(L+1)}, \quad (63)$$

$$\bar{\mathbf{x}}_k \triangleq [\mathbf{x}_k^T, \mathbf{x}_{k-1}^T, \dots, \mathbf{x}_{k-L}^T]^T \in \mathbb{C}^{M(L+1) \times 1}. \quad (64)$$

It implies from (1) and (62) that we should just estimate a compound channel matrix $\bar{\mathbf{H}}_k$ through the use of a compound transmission vector $\bar{\mathbf{x}}_k$, instead of using \mathbf{x}_k . For the proposed OW-RLS algorithm, \mathbf{x}_K is replaced with $\bar{\mathbf{x}}_K$, and Θ_k is replaced with a spatio-temporal and path correlation matrix $\bar{\Theta}_k \triangleq \mathbb{E}[\bar{\mathbf{H}}_j^\dagger \bar{\mathbf{H}}_{j-k}]$. Here, the correlation $\bar{\Theta}_k$ includes any cross-correlation relationship between $\mathbf{H}_{i,p}$ and $\mathbf{H}_{j,q}$ for all symbols i, j and paths p, q . Correspondingly, the extension of the proposed scheme to frequency-selective fading channels does not involve any difficulty. It is noted that, the computational complexity increases because of the matrix extension from \mathbf{H}_K to $\bar{\mathbf{H}}_K$. In the complexity analysis discussed in Table I, we should substitute $M(L+1)$ for M in frequency-selective scenarios of a $(L+1)$ -path model.

It is known that the delayed waves arrived at fractionally sample-spaced timing may occur a larger path correlation between the p -th path channel $\mathbf{H}_{k,p}$ and the q -th path channel $\mathbf{H}_{k,q}$ for any $p \neq q$ after low-pass filtering such as a root-square roll-off Nyquist [43], even when the actual delayed waves are uncorrelated. To combat this influence, we generally use a oversampling approach [44, 45] and a subspace approach [46, 47] for channel estimations. We need more modifications to extend the OW-RLS to these approaches. However, it should be noted that, because the OW-RLS can exploit the correlation information to improve the MSE performance as evaluated in Fig. 10, a certain degree of the performance improvement is expected by the OW-RLS in the presence of such a path

correlation even if we do not employ neither the oversampling nor the subspace tracking.

VI. SUMMARY

This paper investigated on an accurate channel estimation scheme for fast fading MIMO channels including antenna correlations. We theoretically derived an optimum-weighted LS channel estimation method. In the derivation, we clarified that the higher-order LS estimation is not advantageous for optimum-weighting. Moreover, we proposed the OW-RLS algorithm for decision-directed channel tracking, in order to reduce the computational complexity. Through performance evaluations, we verified that the proposed scheme can offer an excellent performance in fast fading channels. Although the proposed scheme requires accurate channel statistics and high-complexity operations, the advantage in the MSE performance may be significantly useful for future wireless communications which allow high mobility and high-speed transmissions.

ACKNOWLEDGMENT

This work is partly supported by JSPS Postdoctoral Fellowships for Research Abroad, TAF Fellowship, and SCAT Research Grant, Japan.

REFERENCES

- [1] T. Koike Akino, "Optimum-weighted RLS channel estimation for time-varying fast fading MIMO channels," in *Proc. IEEE ICC 2007*, Glasgow, Scotland, June 2007.
- [2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.
- [3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–355, Mar. 1998.
- [4] Recommendation ITU-R M. 1645, "Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000."
- [5] H. Taoka, K. Dai, K. Higuchi, and M. Sawahashi, "Field experiments on 2.5-Gbps packet transmission using MLD-based signal detection in MIMO-OFDM broadband packet radio access," in *Proc. WPMC 2006*, San Diego, pp. 234–239, Sept. 2006.
- [6] S. Haykin, *Adaptive Filter Theory, Third Edition*. Prentice-Hall, 1996.
- [7] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. New York: Springer-Verlag, 1990.
- [8] J. Lin, J. G. Proakis, F. Ling, and L. A. Hanoch, "Optimal tracking of time-varying channel: a frequency domain approach for known and new algorithms," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 141–154, Jan. 1995.
- [9] G. V. Moustakides, "Study of the transient phase of the forgetting factor RLS," *IEEE Trans. Signal Processing*, vol. 45, no. 10, Oct. 1997.
- [10] W. Zhuang, "RLS algorithm with variable forgetting factor for decision feedback equalizer over time-variant fading channels," *Wireless Pers. Commun.*, vol. 8, no. 1, pp. 15–29, Aug. 1998.

- [11] Y. Liang, H. Luo, and J. Huang, "Adaptive RLS channel estimation in MIMO-OFDM systems," in *Proc. IEEE ISCT*, vol. 1, pp. 79–82, Oct. 2005.
- [12] Y. Sun, M. Yee, and M. Sandell, "Iterative channel estimation with MIMO MMSE-turbo equalization," in *Proc. IEEE VTC 2003-Fall*, vol. 2, pp. 1278–1282, Oct. 2003.
- [13] C. F. So, S. C. Ng, and S. H. Leung, "Gradient based variable forgetting factor RLS algorithm," *IEEE Trans. Signal Processing*, vol. 83, pp. 1163–1175, 2003.
- [14] S. W. Lee, J. S. Lim, S. J. Baek, and K. M. Sung, "Time-varying signal frequency estimation by VFF Kalman filtering," *IEEE Trans. Signal Processing*, vol. 77, pp. 343–347, 1999.
- [15] K. Kiyoshima, T. Koike, and S. Yoshida, "Algorithm-diversity of MIMO-MMSE with multiple channel estimation schemes for fast fading channels," in *Proc. WPMC 2006*, San Diego, pp. 747–750, Sept. 2006.
- [16] H. Wang and P. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 353–357, May 1996.
- [17] Z. Liu, X. Ma, and G. B. Giannakis, "Space-time coding and Kalman filtering for time-selective fading channels," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 183–186, Feb. 2002.
- [18] L. M. Davis, I. B. Collings, and R. J. Evans, "Extended least squares identification of doubly spread mobile communication channel," in *Proc. Int. Conf. Telecommunications*, Melbourne, Australia, Apr. 1997.
- [19] K. Fukawa and H. Suzuki, "Recursive least-squares adaptive algorithm maximum-likelihood sequence estimation with higher-order state variable model of radio channels—adaptive performance improvement of RLS-MLSE," *IEICE Trans. Commun.*, vol. J75-B-II, no. 7, pp. 415–423, July 1992.
- [20] D. K. Borah and B. D. Hart, "Frequency selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 862–873, June 1999.
- [21] M. Niedzwecki, "Optimization of the window shape in weighted least squares identification of a class of nonstationary systems," *Proc. 7th Con. Anul., Optimizurion of Syst.*, Antibes, France, pp. 889–901, June 1986.
- [22] G. H. Golub and C. F. Van Loan, *Matrix Computations, Third Edition*. Johns Hopkins, 1996.
- [23] W. C. Jakes, *Microwave Mobile Communication*. New York, NY: Wiley, 1974.
- [24] Z. Xu and R. D. Murch, "Performance analysis of maximum likelihood detection in a MIMO antenna system," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 187–191, Feb. 2002.
- [25] J. Benesty, Y. Huang, and J. Chen, "A fast recursive algorithm for optimum sequential signal detection in a BLAST system" *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1722–1730, July 2003.
- [26] K. I. Pedersen, J. B. Andersen, J. P. Kermoal, and P. Mogensen, "A stochastic multiple-input-multiple-output radio channel model for evaluation of space-time coding algorithms," in *Proc. IEEE GLOBECOM*, pp. 893–897, 2000.
- [27] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Select. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.
- [28] H. Özcelik, M. Herdin, W. Weichselberger, J. Wallace, and E. Bonek, "Deficiencies of 'Kronecker' MIMO radio channel model," *Electron. Lett.*, vol. 39, no. 16, pp. 1209–1210, Aug. 2003.
- [29] M. Kiessling, J. Speidel, I. Viering, and M. Reinhardt, "A closed-form bound on correlated MIMO channel capacity," in *Proc. IEEE VTC-Fall*, pp. 859–863, 2002.
- [30] T. S. Rappaport, *Wireless Communications: Principles and Practice, 2nd Ed.* Prentice-Hall, 2002.
- [31] M. D. Yacoub, *Foundations of Mobile Radio Engineering*. CRC Press, Inc. 1993.
- [32] A. Sampath and J. M. Holtzman, "Estimation of maximum Doppler frequency for handoff decisions," in *Proc. IEEE VTC*, pp. 859–862, 1993.
- [33] J. M. Holtzman and A. Sampath, "Adaptive averaging methodology for handoffs in cellular systems," in *Proc. IEEE VTC*, vol. 44, no. 1, pp. 59–66, Feb. 1995.
- [34] H. S. Rad and S. Gazor, "The effect of mobile station rotation on a correlation model for microcellular environments," in *Proc. IEEE GLOBECOM 2004*, July 2004.
- [35] H. S. Rad and S. Gazor, "A cross-correlation model for non-isotropic scattering with non-omnidirectional antennas in MIMO propagation channels," in *Proc. IEEE SPAWC'05*, June 2005.
- [36] M. Dohler, J. Dominguez, and H. Aghvami, "Link capacity analysis for virtual antenna arrays," in *IEEE VTC-Fall*, Sept. 2002.
- [37] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.
- [38] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [39] P. Mitran, H. Ochiai, and V. Tarokh, "Space-time diversity enhancements using collaborative communications," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2041–2057, June 2005.
- [40] T. Miyano, H. Murata, and K. Araki, "Cooperative relaying scheme with space-time code for multihop communications among single antenna terminals," in *Proc. IEEE GLOBECOM*, Dec.–Nov. 2004.
- [41] H. Zhang and G. S. Kuo, "Cooperative diversity for virtual MIMO system in geometry-based stochastic channel model," in *Proc. IEEE ICC*, Glasgow, Scotland, June 2007.
- [42] Z. Wang, X. Ma, and G. B. Giannakis, "OFDM or single-carrier block transmissions?," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 380–394, Mar. 2004.
- [43] J. G. Proakis, *Digital Wireless Communications, Fourth Edition*. McGraw-Hill, 2001.
- [44] X. Dong, Z. Ding, and S. Dasgupta, "Fractional spaced dual channel estimation based on decimated feedback," in *Proc. IEEE WCNC*, pp. 489–494, Mar. 2004.
- [45] X. Wautelet, C. Herzet, A. Dejonghe, J. Louveaux, and L. Vandendorpe, "Comparison of EM-based algorithms for MIMO channel estimation," *IEEE Trans. Commun.*, vol. 55, no. 1, Jan. 2007.
- [46] M. Nicoli and U. Spagnolini, "A subspace method for channel estimation in soft-iterative receivers," in *Proc. Eur. Signal Processing Conf.*, Sept. 2005.
- [47] S. Ferrara, T. Matsumoto, M. Nicoli, and U. Sagnolini, "Soft iterative channel estimation with subspace and rank tracking," *IEEE Signal Processing*, vol. 14, no. 1, Jan. 2007.



Toshiaki Koike Akino received the B.S. degree in electrical and electronics engineering from Kyoto University, Japan, in 2002. He received M.S. and Ph.D. degrees in communications and computer engineering from Kyoto University, in 2003 and 2005, respectively. Since 2006, he has been a postdoctoral researcher in Harvard University. His research interest includes digital signal processing for cooperative multi-user communications. He has been a JSPS Postdoctoral Fellow for Research Abroad since 2008. He received the YRP Encouragement Award 2005 and the 21st TELECOM System Technology Award in Japan.